

JC Rec'd PCT/PTO 01 AUG 2001

09/890463  
PCT/AU00/00056

WO 00/46233

10/PRTS

1

**PIGMENT PROTEIN FROM CORAL TISSUE****Field of the Invention:**

The present invention relates to pigment proteins derived from corals, and particularly polynucleotide molecules encoding the pigment proteins, and uses thereof.

**Background to the Invention:**

There are a number of pigmented and/or fluorescent molecules which have been isolated and characterised from natural sources. Examples include apoequorin, a single polypeptide chain isolated from the luminous jellyfish *Aequorea victoria*, green fluorescent protein (GFP) and renilla luciferase isolated from the Renilla (also called sea pansies) which belong to a class of coelenterates known as anthozoans.

The present applicant has purified pigment protein from various coral tissues (PPCT) and cloned genes encoding PPCT which have unique and useful properties.

**Disclosure of the Invention:**

In a first aspect, the invention provides an isolated polynucleotide molecule comprising a nucleotide sequence encoding a pigment protein from coral tissue (PPCT) capable of emitting fluorescence upon irradiation by incident light, wherein maximal absorbance of said incident light is in the range of 320 - 600 nm (preferably, 550 - 580 nm), and maximal fluorescence emission is in the range 350 - 700 nm (preferably, 400 - 650 nm).

Preferably, the isolated polynucleotide molecule comprises a nucleotide sequence encoding a protein having the N-terminal amino acid sequence:

SVIK (SEQ ID NO:1).

More preferably, the isolated polynucleotide molecule comprises a nucleotide sequence encoding a protein having the N-terminal amino acid sequence:

SVIKQMTYKVYMSGTV (SEQ ID NO: 2).

Still more preferably, the isolated polynucleotide molecule comprises a nucleotide sequence encoding a protein having an amino acid sequence corresponding to the sequence shown hereinafter as SEQ ID NO: 3 or 4.

WO 00/46233

PCT/AU00/00056

2

Even more preferably, the isolated polynucleotide molecule comprises a nucleotide sequence which has at least 80%, more preferably at least 90 % and most preferably at least 95%, identity to the sequence shown hereinafter as SEQ ID NO: 5 or 6.

5 Most preferably, the isolated polynucleotide molecule comprises a nucleotide sequence substantially corresponding to the sequence shown hereinafter as SEQ ID NO: 5 or 6.

10 In a preferred embodiment of the invention, the isolated polynucleotide molecule comprises a nucleotide sequence which is capable of hybridising to the sequence shown as SEQ ID NO: 5 or 6 under high stringency conditions, and is less than 5000 nucleotides in length, although it can be less than 1000 or even 500 nucleotides in length. However, preferably, hybridising polynucleotide molecules according to the invention are at least 18 nucleotides in length.

15 Polynucleotide molecules according to the invention which may hybridise to the sequence shown as SEQ ID NO: 5 or 6 under high stringency conditions, include polynucleotide molecules encoding PPCT isoforms known to the present applicant which are of 221 amino acids in length. Such polynucleotide molecules include a stop codon in place of the codon  
20 encoding Cys<sup>221</sup> shown in SEQ ID NO: 5 or 6.

In a second aspect, the invention provides a protein comprising the N-terminal amino acid sequence:

SVIK (SEQ ID NO:1), said protein being in a substantially purified form.

25 Preferably, the substantially purified protein comprises the N-terminal amino acid sequence:

SVIKQMTYKVYMSGTV (SEQ ID NO: 2).

30 More preferably, the protein comprises an amino acid sequence substantially corresponding to the sequences shown hereinafter as SEQ ID NO 3 or 4.

The protein according to the present invention (PPCT) has the following characteristics:

- i) Aqueous soluble protein (231-235 amino acids in length);
- ii) Absorbance spectra with  $\lambda_{\max}$  ranging from 340-600 nm;
- 35 iii) Light induced;
- iv) Mostly in the form of a dimer, though sometimes trimeric;

WO 00/46233

PCT/AU00/00056

3

v) About 22% amino acid identity and about 56% amino acid similarity with green fluorescent protein (GFP);

vi) Amino acid sequence "QYG" in PPCT instead of "SYG" in GFP chromatophore position;

5 vii) Estimated PI = 9.5; and

viii) Synthesised at high temperatures (e.g. 35°C) and stable at higher temperatures (e.g. 40°C).

10 PPCTs which absorb maximally at 560 - 600 nm are weakly fluorescent (excitation at 560 - 600 nm with emission at 600 - 650 nm) under all conditions. Other PPCTs absorb/excite and emit at different wavelengths. Roughly, there are four classes: Fluorescent behaviours sort PPCTs into those that fluoresce when excited at wavelengths of 300 - 360 nm, 420 - 465 nm, 480 - 490 nm and 550 - 600 nm. These molecules have emission maxima at 400 - 450 nm, 480 - 490 nm, 500 - 505 nm and 610 - 630 nm, respectively  
15 (see Table 4).

PPCT can be purified from the tissues of the following coral families: Pocilloporidae, Acroporidae, Poritidae, Faviidae, Merulinidae, Fungiidae.

20 Preferably, PPCT is purified from the tissue of the coral: *Acropora aspera*, *Acropora digitifera*, *Acropora horrida*, *Acropora formosa*, *Montipora monasteriata*, *Montipora caliculata*, *Pocillopora damicornis*, *Porites murrayensis*, *Porites lobata*, *Plesiastrea versipora* or *Seriatopora hystrix*.

More preferably, PPCT is purified from the tissue of the coral: *Acropora aspera*, *Acropora horrida*, *Montipora monasteriata*, *Montipora caliculata*, *Porites murrayensis*, *Porites lobata*, or *Plesiastrea versipora*.

25 Alternatively, PPCT can be purified from a culture of a recombinant host cell expressing PPCT from an introduced PPCT-encoding polynucleotide molecule.

30 In a third aspect, the invention provides a suitable vector for the replication and/or expression of a polynucleotide molecule according to the first aspect of the present invention.

The vector may be, for example, a plasmid, virus or phage vector provided with an origin of replication, and preferably a promoter for the expression of the polynucleotide and, optionally, a regulator of the promoter. The vector may contain one or more selectable markers, for example an  
35 ampicillin resistance gene in the case of a bacterial plasmid or a neomycin resistance gene for a mammalian expression vector. The vector may be used

WO 00/46233

PCT/AU00/00056

4

*in vitro* for, for example, the production of RNA or, alternatively, for the transfection or transformation of a host cell.

A fourth aspect of the invention provides host cells transfected or transformed with the vector of the third aspect of the present invention.

5        Suitable host cells for cloning or expressing PPCT are prokaryote, yeast, or higher eukaryote cells.

10        Suitable prokaryotes include eubacteria, such as Gram-negative or Gram-positive organisms, for example, *E. coli*, *Bacilli* such as *B. subtilis* or *B. thuringiensis*, *Pseudomonas* species such as *P. aeruginosa*, *Salmonella typhimurium* or *Serratia marcescens*.

15        Eukaryotic microbes such as filamentous fungi or yeast are also suitable hosts for expressing PPCT. *Saccharomyces cerevisiae*, or common baker's yeast, is the most commonly used among lower eukaryotic host microorganisms. However, a number of other genera, species, and strains are commonly available and useful herein, such as *Schizosaccharomyces pombe*; *Kluyveromyces* hosts such as *K. lactis*; filamentous fungi such as *Neurospora* or *Penicillium*; and *Aspergillus* hosts such as *A. nidulans* and *A. niger*.

20        Suitable higher eukaryotic host cells can be cultured vertebrate, invertebrate or plant cells. Insect host cells from species such as *Spodoptera frugiperda*, *Aedes aegypti*, *Aedes albopictus*, *Drosophila melanogaster*, and *Bombyx mori* can be used. Plant cell cultures of cotton, corn, potato, soybean, tomato and tobacco can be utilised as hosts. Typically, plant cells are transfected by incubation with certain strains of the bacterium *Agrobacterium tumefaciens*.

25        Propagation of vertebrate cells in culture (tissue culture) has become a routine procedure in recent years. Examples of useful mammalian host cell lines are monkey kidney CV1 line transformed by SV40 (COS-7, ATCC CRL 1651); human embryonic kidney line (293 or 293 cells subcloned for growth in suspension culture); baby hamster kidney cells (BHK, ATCC CCL 10);  
30        Chinese hamster ovary cells/-DHFR (CHO); mouse sertoli cells, monkey kidney cells (CV1, ATCC CCL 70); human cervical carcinoma cells (HELA, ATCC CCL 2); canine kidney cells (MDCK, ATCC CCL 34), and a human hepatoma cell line (Hep G2). Preferred host cells are human embryonic kidney 293 and Chinese hamster ovary cells.

35        Host cells are transfected or, more preferably, transformed, with expression or cloning vectors of the invention and cultured in conventional

WO 00/46233

PCT/AU00/00056

5

nutrient media modified as appropriate for inducing promoters, selecting transformants, or amplifying the genes encoding the desired sequences. Transformation means introducing DNA into an organism so that the DNA is replicable, either as an extrachromosomal element or by chromosomal integration. Depending on the host cell used, transformation is carried out using standard techniques appropriate to such cells.

In a fifth aspect, the invention provides a process for producing a protein according to the second aspect of the present invention, the process comprises that step of cultivating a host cell transfected or transformed with an (expression) vector of the third aspect of the present invention under conditions suitable for expression of the polynucleotide molecule encoding the protein, and optionally recovering the expressed protein.

Such cells can be used for the production of commercially useful quantities of the encoded protein.

Since PPCT's appear to exhibit stability at high temperatures (e.g. 35 - 40°C), the process of the fifth aspect may be performed at or near optimal cultivating temperatures (e.g. 30 - 37°C).

In a sixth aspect, the invention provides an oligonucleotide probe or primer, the probe or primer having a sequence that hybridises selectively to a polynucleotide molecule according to the first aspect of the present invention.

In a preferred embodiment of the sixth aspect, the oligonucleotide probe or primer comprises at least 8 nucleotides, more preferably at least 18 nucleotides, and most preferably at least 25 nucleotides.

In a further preferred embodiment, the oligonucleotide probe or primer is used as a detectable probe wherein the oligonucleotide is conjugated with a label such as a radioisotope, an enzyme, biotin, a fluorescent molecule or a chemiluminescent molecule.

In a seventh aspect, the invention consists in the use of a protein according to the second aspect as a tissue marker, fluorescent marker or general dyestuff.

It is anticipated that the protein of the invention could be used as a harmless protein-based marker for following gene expression in transformed tissues. The protein of the invention also has potential as a readily manufactured, harmless dyestuff.

WO 00/46233

PCT/AU00/00056

6

In an eighth aspect, the invention provides a sunscreen formulation comprising an effective amount of a protein according to the second aspect, in admixture with a suitable pharmaceutical acceptable carrier or excipient.

5 In a ninth aspect, the invention provides a filter for screening UV or other wavelength(s) of incident light comprising an effective amount of a protein according to the second aspect.

10 Throughout this specification, unless the context requires otherwise, the word "comprise", or variations such as "comprises" or "comprising", will be understood to imply the inclusion of a stated element, integer or step, or group of elements, integers or steps, but not the exclusion of any other element, integer or step, or group of elements, integers or steps.

15 Throughout this specification, where reference is made to percentage levels of nucleotide sequence identity, those levels are as calculated by the BLAST program blastn as described by Altschul, S.F. et al., "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Research, Vol. 25, No. 17, pp 3389-3402 (1997).

20 As used herein, the term "high stringency conditions" refers to conditions that (1) employ low ionic strength and high temperature for washing, for example, 15 mM NaCl/1.5 mM sodium citrate/0.1% NaDodSO<sub>4</sub> at 50°C; (2) employ during hybridisation a denaturing agent such as formamide, for example, 50% (vol/vol) formamide with 0.1% bovine serum albumin, 0.1% Ficoll, 0.1% polyvinylpyrrolidone, 50 mM sodium phosphate buffer at pH 6.5 with 750 mM NaCl, 75 mM sodium citrate at 42°C; or (3) employ 50% formamide, 5 x SSC (750 mM NaCl, 75 mM sodium citrate), 50 mM sodium phosphate (pH 6.8), 0.1% sodium pyrophosphate, 5 x Denhardt's solution, sonicated salmon sperm DNA (50 µg/ml), 0.1% SDS and 10% dextran sulfate at 42°C in 0.2 x SSC (30 mM NaCl, 3 mM sodium citrate) and 0.1% SDS.

30 The term "substantially corresponding" as used herein in relation to nucleotide sequences is intended to encompass minor variations in the nucleotide sequence which due to degeneracy in the DNA code do not result in a change in the encoded protein. Further, this term is intended to encompass other minor variations in the sequence which may be required to enhance expression in a particular system but in which the variations do not result in a change in the functional characteristics of the encoded protein.

35

WO 00/46233

PCT/AU00/00056

7

The term "substantially corresponding" as used herein in relation to amino acid sequences is intended to encompass minor variations in the amino acid sequences which do not result in a change in the functional characteristics of PPCT. These variations may include conservative amino acid substitutions. The substitutions envisaged are:-

G, A, V, I, L, M; D, E; N, Q; S, T; K, R, H; F, Y, W, H; and P, N $\alpha$ -alkylamino acids.

In order that the present invention may be more clearly understood, preferred forms will be described with reference to the following examples and drawings.

**Brief description of the accompanying figures:**

Figure 1 shows a nucleotide sequence of a degenerative primer useful for hybridising to genes encoding PPCT.

Figures 2 provides an electroblot of SDS PAGE of gel filtration HPLC fractions collected for sequencing. The lanes are as follows: 1. *A. horrida*, 2. *M. caliculata*, 3. *M. monasteriata*, 4. *P. lobata*, 5. *P. murrayensis*, 6. *A. formosa*, 7. *S. hystrix*, 8. *P. damicornis*, 9. *S. pistillata*. Molecular weight markers are shown in the unmarked lanes.

Figure 3 shows a amino acid sequence encoded by clone T7SP6BASPOC3 of Figure 3.

Figure 4 shows a amino acid sequence encoded by clone T7SP6BASPOC4 of Figure 4.

Figure 5 shows a cDNA nucleotide sequence of clone T7SP6BASPOC3 encoding PPCT from *A. aspera*.

Figure 6 shows a cDNA nucleotide sequence of clone T7SP6BASPOC4 encoding PPCT from *A. aspera*.

Figure 7 shows a vector suitable for introduction and subsequent bacterial expression of cDNA encoding PPCT.

Figure 8 provides a schematic diagram of molecular models of the principal pigment polypeptide from reef-building corals. A. MOLSCRIPT cartoon, from N-terminus to C-terminus. Arrows represent  $\beta$ -strands and spirals indicate helical segments. heavy atoms of the fluorophore (Q61, Y62 and G63) are shown in ball-and-stick representation. B. Charged, polar and hydrophobic side chains in the immediate vicinity are shown in ball-and-stick representation.

WO 00/46233

PCT/AU00/00056

8

Figure 9 shows the absorbance, excitation and emission spectra of GFP-like pigments (pocilloporin) from A-C. *Acropora aspera* and D-F. *Pocillopora damicornis*. Samples taken at 27-28 min (A. and D.) and 25-26 min (B. and E.) into gel filtration run showing dominant fluorescent excitation (1-3, 5-6) and emission (1-3, 5-6) spectra. Composite absorbance (4, 7, peaks at 580 and 560 nm) and emission spectra (dashed lines) are shown in panels C. and F.

Figure 10 shows the interaction between PPCTs and the Peridinin-chlorophyll-protein (PCP) complex of symbiotic dinoflagellates. Line 1 indicates the excitation spectra of the PCP complex of symbiotic dinoflagellates (*Acropora aspera*), which results in fluorescence emission by the isolated PCP complex at 674 nm (line 2, equivalent to the  $Q_y$  Chlorophyll  $a$  transition). Lines 3, 4, 5 and 6 indicate typical excitation spectra of GFP-like pocilloporins in the reef-building corals *Acropora aspera* and *Pocillopora damicornis*. Corresponding arrows indicate where emission maxima occur. Note the coupling of the emission (arrows) and excitation wavelengths of successive fluorescent pigments. Dotted lines indicate respective excitation peaks for the PCP complex and approximate position of the  $Q_x$  band of Chlorophyll  $a$ . A = Chlorophyll  $a$  (441 nm), B = Peridinin (472 nm) and C = Additional vibrational band of Peridinin (525 nm).

### Example 1:

#### **General Molecular Biology**

Unless otherwise indicated, the recombinant DNA techniques utilised in the present invention are standard procedures, well known to those skilled in the art. Such techniques are described and explained throughout the literature in sources such as, J. Perbal, A Practical Guide to Molecular Cloning, John Wiley and Sons (1984), J. Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbour Laboratory Press (1989), T.A. Brown (editor), Essential Molecular Biology: A Practical Approach, Volumes 1 and 2, IRL Press (1991), D.M. Glover and B.D. Hames (editors), DNA Cloning: A Practical Approach, Volumes 1-4, IRL Press (1995 and 1996), and F.M. Ausubel et al. (Editors), Current Protocols in Molecular Biology, Greene Pub. Associates and Wiley-Interscience (1988, including all updates until present) and are incorporated herein by reference.



WO 00/46233

PCT/AU00/00056

9

**Gene/DNA Isolation**

The DNA encoding a protein may be obtained from any cDNA library prepared from tissue believed to express the gene mRNA and to express it at a detectable level. DNA can also be obtained from a genomic library.

Libraries are screened with probes or analytical tools designed to identify the gene of interest or the protein encoded by it. For cDNA expression libraries, suitable probes include monoclonal or polyclonal antibodies that recognise and specifically bind the protein; oligonucleotides of about 20-80 bases in length that encode known or suspected portions of cDNA from the same or different species; and/or complementary or homologous cDNAs or fragments thereof that encode the same or a hybridising gene. Appropriate probes for screening genomic DNA libraries include, but are not limited to, oligonucleotides; cDNAs or fragments thereof that encode the same or hybridising DNA including expressed sequence tags and the like; and/or homologous genomic DNAs or fragments thereof. Screening the cDNA or genomic library with the selected probe may be conducted using standard procedures as described in chapters 10-12 of Sambrook *et al.*

An alternative means to isolate a gene encoding is to use polymerase chain reaction (PCR) methodology as described in section 14 of Sambrook *et al.* This method requires the use of oligonucleotide probes that will hybridise to the gene.

The oligonucleotide sequences selected as probes should be of sufficient length and sufficiently unambiguous that false positives are minimised. The actual nucleotide sequence(s) is usually based on conserved or highly homologous nucleotide sequences or regions of the gene. The oligonucleotides may be degenerate at one or more positions. The use of degenerate oligonucleotides may be of particular importance where a library is screened from a species in which preferential codon usage in that species is known. The oligonucleotide must be labelled such that it can be detected upon hybridisation to DNA in the library being screened. The preferred method of labelling is to use <sup>32</sup>P-labelled ATP with polynucleotide kinase, as is well known in the art, to radiolabel the oligonucleotide. However, other methods may be used to label the oligonucleotide, including, but not limited to, biotinylation or enzyme labelling.

WO 00/46233

PCT/AU00/00056

10

Nucleic acid having all the protein coding sequence is obtained by screening selected cDNA or genomic libraries, and if necessary, using conventional primer extension procedures as described in section 7.79 of Sambrook *et al.*, to detect precursors and processing intermediates of mRNA that may not have been reverse-transcribed into cDNA.

Another alternative method for obtaining the gene of interest is to chemically synthesise it using one of the methods described in Fingels *et al.* (Agnew Chem. Int. Ed. Engl. 28: 716-734, 1989). These methods include triester, phosphite, phosphoramidite and H-Phosphonate methods, PCR and other autoprimer methods, and oligonucleotide syntheses on solid supports. These methods may be used if the entire nucleic acid sequence of the gene is known, or the sequence of the nucleic acid complementary to the coding strand is available, or alternatively, if the target amino acid sequence is known, one may infer potential nucleic acid sequences using known and preferred coding residues for each amino acid residue.

#### Purification Scheme

- (i) Extraction of coloured pigment from coral tissue by immersing coral overnight at 4°C in 0.6 M Potassium Phosphate buffer pH 6.65.
- (ii) Gel filtration (Pharmacia, Superose HR 10/30) monitored at 280 nm &  $\lambda_{\text{max}}$  eluting with Phosphate buffer (peak collection)
- (iii) Desalt and lyophilise
- (iv) Separate on 15% SDS-PAGE
- (v) Electroblot onto PVDF using CAPS buffer
- (vi) N-terminal Sequence analysis (Australian National University Sequencing Facility)

#### cDNA Library

Species: Blue tipped *Acropora aspera*

- (i) Isolation of mRNA: Qiagen mRNA Kit (poly A+) or, more preferably, by the method described by Chomczynski and Sacchi (Anal. Biochem. 162, 156-159 (1987)).
- (ii) cDNA made with SMART PCR cDNA Synthesis Kit from CLONTECH
- (iii) cDNA ligated into Stratagene LambdaZapII EcoR1 cut/CIAP and packaged by Stratagene Gigapack II.

#### PCR out of Library

- (i) Using gel purified oligomer as 5' primers and SMART PCR adaptor 3' primers - PCR amplified genes

WO 00/46233

PCT/AU00/00056

11

(ii) Gel purified product ligated into pGemT easy vector (Promega)  
LICOR sequencing from T7 and SP6 primers and translated sequence to give  
the amino acid sequence.

#### Production of Mutants

5 It is anticipated that the fluorescence properties of PPCTs which are  
weakly fluorescent may be improved by mutagenesis. Such mutants (and the  
polynucleotide molecules encoding them) are to be considered as part of the  
present invention.

10 The sites for mutation can be modified individually or in series, for  
example by:

- (i) substituting a target residue first with conservative amino acid choices  
and then with more radical selections depending upon the result to be  
achieved,
- (ii) deleting a target residue, and/or
- 15 (iii) inserting residues or other ligands adjacent to the target residue.

One useful method for identification of residues or regions for  
mutagenesis is called "alanine scanning mutagenesis" as described by  
Cunningham and Wells (*Science* (1989) 244: 1081-1085). Here, a target  
residue or group of target residues are identified (e.g. charged residues such  
20 as Arg, Asp, His, Lys and Glu) and substituted by a neutral or negatively  
charged amino acid (most preferably, Ala or poly-Ala) to affect the  
interaction of the amino acids with the surrounding aqueous environment in  
or outside the cell. Those domains demonstrating functional sensitivity to  
the substitutions then are refined by introducing further or other mutations.  
25 Thus, while the site for introducing an amino acid sequence mutation is  
predetermined, the nature of the mutation *per se* need not be predetermined.  
For example, to optimise the performance of a mutation at a given site,  
alanine scanning or random mutagenesis may be conducted at the target  
codon or region and the expressed mutants are screened for the optimal  
30 combination of desired function.

#### Example 2:

##### **METHODS**

##### **Isolation and purification of pigment proteins**

35 Pigment proteins were purified and electrophoresed under denaturing  
conditions using the method described in Dove et al., *Biol Bull*, 189, 288-297,

WO 00/46233

PCT/AU00/00056

12

1995. The resulting SDS-PAGE subunits were transferred onto PVDF paper in the presence of CAPS buffer. The 28 kD subunits isolated from five species of hermatypic corals were then sequenced by the Biomolecular Resource Facility at the Australian National University.

5 **RNA isolation - cDNA library construction - thermocycler conditions**

Branch tips (blue) of *Acropora aspera* were collected at noon on November 24 1997 from the Heron Island reef flat and taken immediately to the research station. The total RNA from 10 tips (approximately 1 cm long) were isolated immediately and a cDNA library was made using a SMART  
10 PCR cDNA Synthesis Kit from Clontech Laboratories (Palo Alto, CA, USA). Full length mRNA sequences of PPCT were amplified out of the library using gel purified degenerative primers made to the first 10 amino acids at the 5' end or N-terminus of the most common sequence (Table 1) and SMART PCR CDS 3' primers. This yielded a single product that was approximately 760 bp  
15 in length. Gel purified product was ligated into pGemT easy vector (Promega, Madison, WI, USA) and sequenced using T7 and SP6 primers (2 clones, forward and reverse directions, LICOR, USA). The resultant derived proteins had 231 amino acids (POC3) and 235 amino acids (POC4) with only a few functionally similar amino acid substitutions.

20 **Model construction**

Sequence alignment was carried out using MALIGN (Johnson et al. 1993) with the default scoring matrix derived from multiple-structure alignments and visually edited to optimise the alignment of conserved and chemically similar residues. The program MODELLER was used to generate  
25 the POC4 model, based on the template structure of the green fluorescent protein (PDB 1GFL). The initial model was iteratively refined by in-built molecular dynamics with simulated annealing protocols, to improve the structural quality as computed by PROCHECK. GRASP was used to build the molecular surface of the protein model.

30 **Measurement of Emission and Excitation spectra**

The following gel filtration fractions were analysed for excitation and emission spectra: (i) Gel filtration fractions at intervals of one minute were collected between 25 min and 29 min that correspond approximately to trimeric through monomeric proteins of subunits ranging between 25 kD and  
35 29 kD, (ii) Narrow gel filtration fractions collected around the wavelength of maximal absorbance for protein purification, these ranged from

WO 00/46233

PCT/AU00/00056

approximately 26 min through approximately 28 minutes depending on coral species (Table 1).

Isolation of the peridinin-chlorophyll-protein (PCP complex from zooxanthellae. The water soluble peridinin-chlorophyll-protein (PCP) complex was isolated from zooxanthellae isolated from several corals (*Montipora digitata*, *Acropora aspera*) by homogenizing tissue removed from corals by using an air-brush. A symbiotic dinoflagellate rich fraction was created by spinning the homogenate at 3000 rpm for 2 min. The pellet was incubated with ice-cold phosphate buffer. PCP complexes were further purified and electrophoresed under denaturing conditions using the method described in Dove et al 1995 *supra* except that fractions were collected between 30-31 min.

### RESULTS AND DISCUSSION

PPCT from ten species of reef-building corals were isolated. The isolated proteins had native molecular weights (approximate) that ranged from 42 to 69 kD (Table 1) and consisted of subunits that were approximately 28 kD in size (Figure 2). The N-termini of subunits from 6 species of Acroporid and Poritid corals were sequenced. Although there were a number of background sequences in the 28 kD band, the sequence beginning with the amino acids SVIAK was common to all families (Table 1). Ion exchange chromatography using a Mono-Q column (Pharmacia, Sweden) and a variety of pH gradients and salt concentrations failed to give further resolution of the gel filtration fraction. This suggests that the surface charges on these polypeptides are very similar. In order to derive the complete sequence of the polypeptide with the common N-terminal sequence, cDNA libraries were constructed from a blue pigmented coral, *Acropora aspera*. This resulted in the successful amplification of the full length mRNA coding this protein. Amplification resulted in two clones with a few nucleic acid differences, yielding proteins of length 231 (POC3) (see Figure 5) and 235 (POC4) (see Figure 6), both having isoforms with stop codons at amino acid 221. Database searching matched these sequences with Green Fluorescent Protein (GFP) from the jellyfish *Aequorea victoria* (Cl. Scyphozoa). Significantly, GFP is also a dimer with a 28 kd subunit (Tsien, *Ann. Rev. Biochem.*, 67, 509-544, 1998). The alignment of POC4 with GFP and the GFP-like proteins isolated from *Discosoma* sp. (Order Corallomorpharia, Cl. Anthozoa, Matz et al., *Nature Biotech.* 17, 969-973, 1999) showed 19.6% identity with

WO 00/46233

PCT/AU00/00056

14

GFP and 65.3% identity with drFP583 (Table 2). When concentrated, the narrow gel filtration fractions have approximately the same coloration as the coral from which they were derived (i.e. blue, purple or pink). The pigments observed in corals under visible light are therefore due to this family of GFP-proteins. A dimer with a 28 kD subunit matches the profile of GFP.

The construction of a 3-dimensional model revealed that the PPCT sequence can adopt the 'β-can' structure of green fluorescent protein (PDB 1GFL) (Yang et al., Nature Biotech, 14, 1246-1251, 1996), with 11 β-strands and a chromophore (composed of the modified residues Gln-61, Tyr-62 and Gly-63) within the protein (Figure 8B). The model shows that the chromophore is formed by the "QYG" amino acid motif. The absence of part of the N-terminal cap (see alignment in Table 2, Figure 8A,B) in the PPCT suggests that this molecule may have a reduced ability to fluoresce as this region was found to be crucial for GFP in N-terminal truncation experiments. It is suspected that this sequence represents the high absorbance-low emission form of the pigment discussed below (excitation maxima at 570 nm, Figure 9). The GRASP molecular surface of the PPCT model (not shown) reveals no channels, suggesting that the fluorophore is well protected from solvent. The immediate vicinity of the fluorophore appears completely shielded from bulk solvent, with 26 residues within 5 Å (listed in Table 4 and shown in Figure 8A,B) as compared to 19 residues in GFP. This increase is partially attributable to the hydrogen bonds formed by the side chain of Gln-61. Two resonance forms have been proposed for the fluorophore, one with a partial negative charge either on the benzyl oxygen (OH) of Tyr-62 and the other with a negative charge on the carbonyl oxygen of Tyr-62 (part of the imidazolidone ring). Both resonance forms appear equally stabilized by Arg-192 and Arg-90 (Table 4, in bold), respectively and it is thus not possible to speculate on which form is dominant from the model. Yang et al. (1996) *supra* have suggested that mutation of Thr-203, Glu-222 or Ile-167 of GFP would directly affect the fluorophore, from steric and electrostatic considerations. The corresponding residues in PPCT, Arg-192, Glu-210 and Met-158, are all located in the vicinity of the fluorophore, suggesting that these are critical residues for the function of the protein. Of special importance is the presence of tryptophan residues within the "β-can" structure: Trp-88 (corresponding to Asn-94 of GFP), within van der Waals contact of the fluorophore, with Trp-53 and Trp-138 (10.8 Å and 6.7 Å,

WO 00/46233

PCT/AU00/00056

15

respectively) from the fluorophore. The latter, aligning with Trp-57 (13 Å in WT; Yang et al. (1996) *supra*) and Tyr-145 of GFP, possibly contribute to charge delocalization. It is suggested that these tryptophans are important in stabilizing and protecting the PPCT from ultraviolet radiation, which adds to its potential function under high levels of solar radiation (see below). The model structure also explains the unusual stability of PPCTs, that are expressed in shallow water environments that may exceed 35° C as opposed to GFP with a maximum at 25° C.

Examination of the fluorescence of native PPCT from ten species of reef-building corals revealed a broad range of fluorescent behaviour which was similar to that seen in both native and mutant GFPs (Tsien et al., Biochem., 67, 509-544, 1998). Fluorescence behaviours roughly sort molecules into those that fluoresce when excited at wavelengths of 300-360 nm, 420-465 nm, 480-490 nm and 550-600 nm (Table 4). These molecules had emission maxima at 400-450 nm, 480-490 nm, 500-505 nm and 610-630 nm respectively. The group of molecules that absorb in the 550-600 nm range and emit in the 610-630 nm range have very low intensity fluorescence as can be seen from the high absorbance reading in this range for narrow gel filtration fractions (and high protein content), and their comparatively low fluorescence (Figure 9). There were also quite substantial differences between species with respect to the types of PPCTs that were present in each coral species. PPCT derived from the blue branch tips of the Acroporid coral *Acropora aspera*, for example, fluoresced under both UVA and visible light and showed three distinct excitation maxima of 340 nm, 420 nm and 576 nm (Figure 9 A,B). Excitation at these wavelengths led to fluorescent emissions at 445 nm (blue), 490 nm (blue-green) and 630 nm (orange-red, weak) respectively. If HPLC samples were taken at 25-26 min into a run, PPCT with an excitation maximum of 340 nm dominated while PPCTs with excitation maxima of 420 nm and 576 nm dominated samples taken at 26-27 min. Several types of PPCT were found within the tissues of *Acropora aspera*, with some forms having a greater potential to form homo, possibly hetero, dimers or trimers than others. The different elution times were also seen with PPCTs from all other species. GFP and mutants have single (major, excluding shoulders) excitation and emission maxima, further strengthening the case that each pair of excitation and emission spectra found represent individual PPCTs.

WO 00/46233

PCT/AU00/00056

16

The bright colours of reef-building corals are consequently due to a combination of the absorptive and fluorescent properties of one or more GFP-like pocilloporins. For example, the blue colour of the branch tips of *Acropora aspera* is due to strong absorbance in the green to orange sections of visible spectra (peak absorbance = 580 nm, and no absorbance at wavelength longer than 650 nm) and substantial blue fluorescence emission (Figure 9C). In purple species, the absorbance maxima shift slightly away from the red end of the spectra (Table 3), but strong blue fluorescence at excitations of approximately 340 nm and/or 420-465 nm is retained. Pink coral species like *Pocillopora damicornis* (Figure 9F), on the other hand, have PPCTs that have peak absorbance at 560 nm and no absorbance beyond 600 nm. In this case, mostly yellow and red light is transmitted and reflected back by the underlying white coral skeleton and blue fluorescence arising from excitation at 358 nm or 484 nm is relatively weak or is non-existent. Many corals are highly fluorescent at 340 nm but either do not have any visible host colour under visible light or are a luminescent green. These corals probably have the 340 nm and 420-465 nm excitation variety of GFP-like protein (green emitters) and do not have the high absorbance-low emission blue or pink varieties of PPCT described here.

Green Fluorescent Protein was discovered in the bioluminescent jellyfish (Cl. Scyphozoa) *Aequorea victoria* where it has an excitation maximum at 400 nm (shoulder at 475 nm), and an emission peak at 509 nm. In this species, GFP acts to emit green light after absorbing the blue-green light originating from the  $\text{Ca}^{2+}$  activated photoprotein, aequorin (Morin et al., J. Cell Physiol., 77, 313-318). Reef-building corals are not bioluminescent. So, why do reef-building corals have fluorescent GFP-like proteins like PPCTs? The molecular properties of PPCT suggest that these proteins function in photoprotection and light harvesting as has been suggested by several authors. Different PPCTs appear to have different roles. The high absorbance-low emission varieties specifically absorb light from 550-600 nm at wavelengths which is too long for absorption by the PCP complex (410-550 nm, Figure 13) and emit at wavelengths too short for the  $Q_y$  band of chlorophylls *a* and *c*. The absorbance range of these low emitting PPCTs coincides with the  $Q_x$  bands of chlorophylls *a* and *c*. The  $Q_x$  band is poised to absorb scattered light due to the different dipole orientations of these chlorophylls. As the proportion of scattered light to total irradiance increases



WO 00/46233

PCT/AU00/00056

17

with light intensity, this particular PPCT is likely to act as a light-dependent screening pigment and reduce photoinhibition of the symbiotic dinoflagellates in high light habitats. The fact that the concentrations of the high absorbance-low emission PPCTs are correlated with PAR and not UV is further evidence of this role.

Other PPCTs convert UV radiation (down to at least 300 nm, Figure 9a) and blue light into longer wavelengths. This suggests two additional roles of PPCT under both high and low light conditions respectively. The UV component of solar radiation has been shown to have a major and negative effect on coral and their dinoflagellate symbionts. By converting UV radiation into blue light (440-445 nm) and then (in many cases) into blue-green light (480-490 nm and 500-505 nm), UV radiation can be converted into less harmful visible radiation that can be absorbed by the PCP complex of the symbiotic dinoflagellates (Figure 10). The PCP complex is a highly regulated light-harvesting complex that can either direct the captured energy toward the two photosystems or dump excess light energy as heat via transfers to associated carotenoids. It is striking how some of these proteins direct energy away from the Soret band at about 430 nm and into a trough in the absorbance wavelengths of the PCP complex (Figure 10). By doing this, they clearly have a role in limiting the direct entry of light into Chlorophyll *a* by converting some light into that that can be absorbed by other components of the PCP complex. The behaviour of some PPCTs in converting UV light to longer wavelengths also appears poised to play a role in low light habitats. By converting non-photosynthetic UV light into blue and green light, PPCT extends the range of wavelengths over which symbiont photosynthesis can occur. Consequently, in low light habitats, the 300-360 nm and 420-465 nm forms of PPCT can function as a host based accessory pigments that increase light availability for reef-building corals that grow there. The observation that fluorescent pigment granules enhance the photosynthetic rate of deepwater corals further supports this additional role of PPCT.

WO 00/46233

PCT/AU00/00056

18

**Table 1. Size and N-terminus amino acid sequences of subunits from HPLC gel filtration fractions.**

Coral Species	Elution time (min)	Approximate native molecular weight (Kd)	N-terminal amino acid sequences of resulting subunit (28 kD)
<i>Stylophora pistillata</i>	27.9	42	Nd
<i>Pocillopora damicornis</i>	27.6	46	Nd
<i>Seriatopora hystrix</i>	26.9	57	Nd
<i>Acropora digitifera</i>	27.6	46	Nd
<i>Acropora horrida</i>	26.5	64	SVIAKQMTYKVYMSGTV (SEQ ID NO: 2)
<i>Acropora formosa</i>	26.5	64	Nd
<i>Acropora aspera</i>	26.4	65	SVIAKQMTYKVYMSGTVN* (SEQ ID NO: 7)
<i>Montipora monasteriata</i>	26.4	65	SVIAK (SEQ ID NO: 1)
<i>Montipora caliculata</i>	26.2	69	SVIAKQMTYKVYMSGTVN (SEQ ID NO: 7)
<i>Porites murrayensis</i>	26.3	67	SVIAKQMTYKVYMSGTVN (SEQ ID NO: 7)
<i>Porites lobata</i>	26.3	67	SVIAKQMTYKVYMSGTVNNHYEFVT (SEQ ID NO: 8)

5

Asterisk means sequence derived from cDNA

WO 00/46233

PCT/AU00/00056

19

**Table 2. Nucleotide and amino acid sequence of mRNA encoding a GFP-like protein from the reef-building coral *Acropora aspera*.**

poc4	QVAKQSTKRYEETNGHFEEDKCKPYEETRTTKGGPL	
drfp385	MRSSKNVKEFARKREETNGHFEEDKCKPYEETRTTKGGPL	
gfp	MSKGEELTGVPPLLEEDGONCHSGSEEDATYKTEKITTGKL	
Consensus	.....	
poc4	PPSDQYGSFKPEDI--PKKPEGTWERIMNEDGAVETSDSQQGNC	
drfp385	PPSDQYGSFKPADI--PKKPEGTWERIMNEDGAVETSDSQQGNC	
gfp	SVGSSVQGSFKPDHMKRHEKAPPEGVQGETFKDDGNKKRSEKEDT	
Consensus	.....	
poc4	IHKKSLNPPNGPT-QKDTGWEPTRFA--RDGMINFTAKLEGCG-HY	
drfp385	IHKKSLNPPSDGPT-QKDTGWEPTRFA--RDGMINFTAKLEGCG-HY	
gfp	VERIEKKIDKEDGNKGHKSENYNSNYYMADKQKNGKNNKRNIEDGSVQL	
Consensus	.....	
poc4	CHFSYKAKK-PVRMPGYDYDKIDTNNHKDYTS-VEGCEISIRKPVVAWCFFRV	(SEQ ID NO: 4)
drfp385	VFYSYMAKK-FVQLPGYDYDKIDTSHNEDYTI-VEGYERERHHLFL	(SEQ ID NO: 9)
gfp	DYFCQTPIGDGPVLLPDNYESQALSKDPNEKRDHIVLEFETAGIT	(SEQ ID NO: 10)
Consensus	.....	

WO 00/46233

PCT/AU00/00056

20

**Table 3. Maximal absorbance, and excitation and corresponding emission peaks of GFP-like proteins in reef-building corals.**

Coral Species	Colour	Absorbance (nm)	Excitation (nm)	Emission (nm)
<i>Pocillopora damicornis</i>	Pink	560	358 484 570	440* 499* 625
<i>Seriatopora hystrix</i>	Pink	560	462	482
<i>Acropora digitifera</i> +	Purple-blue	578 (550 shoulder)	425 570	490 590
<i>Acropora horrida</i> +	Blue	579	345 420 575	400 485 625
<i>Acropora aspera</i>	Blue	580 (550 shoulder)	340 420-450 480 576	445* 490* 500 630*
<i>Montipora monasteriata</i> +	Purple	574-578	420-450 570	490 610
<i>Montipora caliculata</i> +	Purple	579 (545 shoulder)	420-450-465	485
<i>Porites lobata</i>		576 (545 shoulder)		
<i>Porites murrayensis</i> +	Purple	576 (545 shoulder)	420 530 570	485 550 625
<i>Plesiastrea versipora</i>	Green	492	492	505

\* indicate pigments profiles that are shown in Figure 9.

+ means same gel filtration fraction as that used in preparation for sequencing.

WO 00/46233

PCT/AU00/00056

21

**Table 4. Residues within 5 Å of the fluorophore, with details of atomic contacts. Atom numbers are in accordance with PDB nomenclature.**

Interactions stabilizing the resonance forms of the fluorophore are in bold.

<u>Protein</u>		<u>Fluorophore</u>		<u>Distance</u>
Residue	Atom	Residue	Atom	<sup>a</sup> (Å)
Glu-210	OE2	Gln-61	OE1	2.5
Ser-64	N	Gly-63	O	2.6
Arg-90	NH2	Tyr-62	O	2.7
Arg-192	NH1	Tyr-62	OH	2.7
Glu-210	OE1	Gln-61	NE2	2.7
Ser-57	O	Gln-61	N	2.9
Gln-37	OE1	Gln-61	CG	3.0
Pro-58	O	Gly-63	N	3.0
Val-39	CG1	Gln-61	OE1	3.3
Glu-143	OE1	Tyr-62	CD2	3.4
Tyr-115	CD2	Gly-63	O	3.5
Ser-106	OG	Gly-63	O	3.6
Asn-141	CB	Tyr-62	OH	3.7
Leu-194	CD2	Tyr-62	CE1	3.7
Tyr-86	OH	Gly-63	O	4.2
Met-158	CE	Tyr-62	OH	4.2
Gln-208	CB	Gln-61	NE2	4.3
Glu-139	O	Tyr-62	OH	4.5
Asn-156	OD1	Tyr-62	CE2	4.6
Glu-171	O	Tyr-62	CD2	4.7
Leu-56	O	Gln-61	N	4.8
Trp-88	NE1	Tyr-62	O	4.8
Cys-209	O	Gln-61	NE2	5.0

WO 00/46233

PCT/AU00/00056

22

**Example 3:**

POC3 and POC4 were PCR amplified out of the pGEM-EASY cloning vector using the following forward and reverse primers:

5        POC (FORWARD): 5'-TCC GTT ATC GCT AAA CAG AT-3'  
         (SEQ ID NO: 11)  
         POC3(231) (REVERSE): 5'-TTT GTG CCT TGA TTT GAC TC-3'  
         (SEQ ID NO: 12)  
         POC4(235) (REVERSE): 5'-CGC CAC TGC GTA TTT GTG CC-3'  
10        (SEQ ID NO: 13)  
         POC4(220) (REVERSE): 5'-GGC GACCAC AGG TTT GCG TG-3'  
         (SEQ ID NO: 14)

15        Amplified sequences were gel purified and inserted into the bacterial  
pBAD TOPO expression vector using the pBAD TOPO TA cloning kit  
(Invitrogen, Carlsbad, CA, USA), prior to transformation into One Shot  
Competent Cells (Invitrogen, Carlsbad, CA, USA). Reading frame and  
occurrence of stop codon at position AA221 in longer sequences (231 amino  
20        acids and 235 amino acids) were checked by ABI nucleotide sequencing from  
pBAD forward and reverse priming sites in the pBAD TOPO expression  
vector.

25        Protein expression from pBAD TOPO was regulated by optimising  
Arabinose concentration. Expressed proteins are purified by: (i) binding the  
6 x His tag that is synthesised at the C-terminus of proteins engineered to use  
the pBAD TOPO transcription stop codon, and (ii) by gel filtration  
30        chromatography for expressed proteins lacking the 6 x His tag.

**Example 4:**

30        The ability of PPCTs to absorb light in the range of 320 - 700 nm,  
which covers UVA and UVB radiation makes them useful screening  
components of sunscreen formulations. Such formulations would include,  
for example, an effective screening amount of PPCT, optionally in admixture  
with a pharmaceutically acceptable carrier or excipient. Sunscreen  
formulations including PPCT may be produced by forming an emulsion by  
35        mixing PPCT with an agent or composition which renders the PPCT non-  
water soluble. Alternatively, the PPCT may be synthesised with a "tag"

WO 00/46233

PCT/AU00/00056

23

peptide or compound such as a long chain fatty acid (e.g. by linking the "tag" to, or near, the C-terminus of the PPCT), to render the PPCT non-water soluble and thereby capable of forming an emulsion in an aqueous carrier or excipient. Methods of acylating proteins at, for example, carboxy or amino groups (e.g.  $\epsilon$ -amino groups) of proteins are well known in the art. The amount of PPCT present in a sunscreen formulation would generally be at least 1% by weight of the formulation. A preferred sunscreen formulation including PPCT (acylated with a fatty acid with greater than 10 carbon atoms) comprises:

- (i) about 3% cetostearyl alcohol, PEG-40 castor oil and/or sodium cetearyl sulphate,
- (ii) about 10% decyl oleate,
- (iii) about 1.5% sodium hydroxide (45% solution),
- (iv) about 0.1 to 0.5% disodium edetate,
- (v) about 0.5% carbomer,
- (vi) about 1 to 10% acylated PPCT, and
- (vii) the balance being water.

The sunscreen formulation may also include other screening agents well known in the art, for example, propyl hydroxybenzoate, dimethylaminobenzoate (PABA), phenyl salicylates, and/or octyl methoxycinnamate.

#### Example 5:

The ability of PPCTs to absorb light in the range of 320 - 700 nm also makes them useful for incorporation into various materials such as glass and polymeric materials such as perspex for use as filters for application in, for example, sunglass lenses, cameras, aquariums and greenhouses.

The absorbance range of PPCTs of 320 - 700 nm covers the solet band (430 nm) of chlorophyll a, but more specifically the  $Q_x$  band of chlorophyll a and c (550-600nm). The  $Q_x$  band specifically absorbs scattered light associated with high light intensities. PPCTs incorporated into filters may therefore function to reduce photoinhibition of plants in high light environments. Further the ability of PPCTs to emit light at a wavelength higher than their excitation or maximal absorbance wavelength provides the possibility of transmitting across, for example, glass, the energy associated with light wavelengths which would not otherwise be transmitted. Where

WO 00/46233

PCT/AU00/00056

24

that energy is transferred at a wavelength in the range of photosynthetically active radiation (PAR), that is 400 - 550 nm, the use of PPCT in filters offers the possibility of generating improved growing conditions for plants, algae, corals etc. For example, use of the PPCT from *A. horrida* which is excited by light of 320 - 360 nm, would allow the absorbance of UVA light (which is not greatly transmitted by glass) to be retransmitted at the PAR wavelength of 430 - 450 nm.

The high-temperature tolerance of PPCT ( $>40^{\circ}\text{C}$ ) enables PPCT to be readily included into, or coated onto, glass and other polymeric materials (especially polymeric materials with relatively low melting temperatures). The amount of PPCT included in glass and other polymeric materials, or alternatively included into a coating material, would generally be greater than 1 % by weight.



WO 00/46233

PCT/AU00/00056

25

5 It will be appreciated by persons skilled in the art that numerous variations and/or modifications may be made to the invention as shown in the specific embodiments without departing from the spirit or scope of the invention as broadly described. The present embodiments are, therefore, to be considered in all respects as illustrative and not restrictive.

PCT/AU00/00056 WO 00/46233